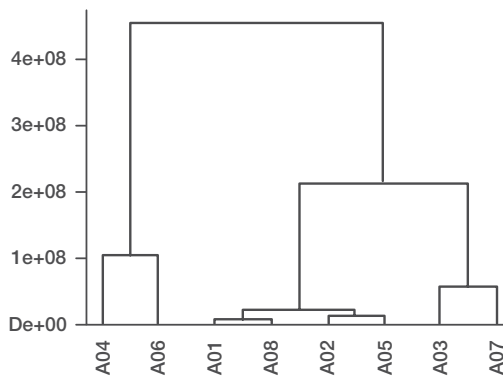


Figure 3: Cluster Dendrogram



Clustering dendrogram of the *Listeria* samples based on SNP calling results. The clustering closely matches the results obtained with PFGE genotyping.

Two of the samples sequenced on the MiSeq were identical to two samples that the CDC sequenced on the Genome Analyzer. Only the N50 data of those four samples was compared.

MiSeq *de novo* sequencing results were concordant with the data generated by the Genome Analyzer (Table 1). The Genome Analyzer data were from a 1 × 70 bp run and the MiSeq data from a 2 × 150 run. To perform a meaningful comparison, both data sets were downsampled to the same coverage. Several assemblies were performed over a range of k-mer sizes, before choosing those with the highest N50 (50% of the assembled genome lies in contigs of this size or larger) (Table 2). MiSeq and its paired-end data provided a larger N50 scaffold size, demonstrating the value of a paired-end run for *de novo* sequencing. All assemblies had a high agreement with the reference genome.

Figure 2 shows the cumulative scaffold sizes for two MiSeq samples and two corresponding Genome Analyzer data sets. Lines with higher slope represent assemblies which are more contiguous. Both MiSeq assemblies are more contiguous and therefore have higher slopes.

Resequencing of the outbreak samples yielded > 92% alignment with the reference genome. Variant analysis filtered for $Q(\text{SNP}) / Q(\text{indel}) > 20$ demonstrated samples 1, 2, 5, and 8 were closely related, with two outlier groups formed by samples 4 and 6, and 3 and 7. The digital NGS data enabled the creation of a dendrogram to illustrate the hierarchical clustering of the samples (Figure 3). The results matched what PulseNet obtained viewing the analog results (bands) generated by PFGE, and were supported by the raw SNP and indel calls, and the analysis of missed/extra segments obtained in *de novo* assembly.

Illumina, Inc. • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

© 2012, 2014 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPPro, DASL, DesignStudio, Eco, GALix, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.

Pub. No. 770-2012-021 Current as of 10 November 2014

Summary

NGS has the potential to become a valuable, cost-effective tool for PulseNet's laboratory network, supporting its mission of public health surveillance, outbreak detection, identification, source attribution, and containment. This retrospective analysis demonstrates the potential of the MiSeq platform to generate *de novo* and resequencing data in an outbreak scenario, generating results that are comparable with the CDC Core Facility's Genome Analyzer. The MiSeq system's fast turnaround matches the speed of PFGE, while its higher resolution and accuracy enhance the ability to identify closely related bacterial isolates. Its one-touch operation supports its use in regional laboratories where ease of use is essential. Data from the MiSeq and other Illumina sequencers can also be analyzed in a cloud computing environment, such as Illumina BaseSpace™, where it can be shared in real-time with regional, national, and international public health agencies during an outbreak.

References

1. <http://www.cdc.gov/listeria/outbreaks/cantaloupes-jensen-farms/index.html>
2. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, et al. (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 364: 730–739.
3. <http://www.sanger.ac.uk/about/press/2012/120613.html>
4. http://www.broadinstitute.org/annotation/genome/listeria_group/Genome-Descriptions.html#LM_10403S_V6